A communication-efficient nonlocal measurement with application to communication complexity and bipartite gate capacities

Aram W. Harrow[1] and Debbie W. Leung[2]

[1] *Department of Mathematics, University of Bristol, Bristol, BS8 1TW, United Kingdom. a.harrow@bris.ac.uk*

[2] *Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada. wcleung@iqc.ca*

(Dated: October 21, 2009)

Two dual questions in quantum information theory are to determine the communication cost of simulating a bipartite unitary gate, and to determine their communication capacities. We present a bipartite unitary gate with two surprising properties: 1) simulating it with the assistance of unlimited EPR pairs requires far more communication than with a better choice of entangled state, and 2) its communication capacity is far lower than its capacity to create entanglement. This suggests that 1) unlimited EPR pairs are not the most general model of entanglement assistance for two-party communication tasks, and 2) the entangling and communicating abilities of a unitary interaction can vary nearly independently. The technical contribution behind these results is a communication-efficient protocol for measuring whether an unknown shared state lies in a specified rank-one subspace or its orthogonal complement.

Introduction. Many basic questions in quantum information theory can be phrased as determining the rates at which standard communication resources (EPR pairs, noiseless qubit channels, etc.) can be converted to and from more specialized resources (such as an available noisy channel, or computation of functions of interest with distributed inputs). Typically local operations are allowed for free; sometimes entanglement is as well. For example, channel capacities are the maximum rates at which noisy channels can be turned into noiseless ones, while the quantum communication complexity of a function f is related to the minimum rate at which noiseless quantum communication is turned into evaluations of f.

In quantum mechanics, the most general interaction between two systems, given sufficient isolation from the environment, is a bipartite unitary quantum gate U. We will think of the systems (A and B) as each comprising n qubits, and as being held by two parties, Alice and Bob.

A fundamental goal of quantum information processing is to simulate interactions (i.e. unitaries) using as few resources as possible. This Letter investigates these simulation costs when different types of entanglement are given for free. We will define $C^{\mathrm{ent}}_{\mathrm{sim},\epsilon}(U)$ to be the number of bits of classical communication necessary to simulate U up to error ϵ if Alice and Bob are allowed to start with an entangled state of their choice. (Given free entanglement, the quantum and classical communication costs differ by a factor of exactly 2, due to teleportation [1] and super-dense coding [2].) The canonical form of entanglement is the EPR pair, since it can be converted to many copies of any other state using an asymptotically vanishing amount of communication per copy [3]. Accordingly, we also let $C^{\mathrm{EPR}}_{\mathrm{sim},\epsilon}(U)$ denote the classical communication cost of simulating U up to error ϵ given unlimited EPR pairs.

Also of interest is the effectiveness of unitaries at sending classical messages or generating entanglement. The ultimate limit to which this can be done is given by the rate achievable with an asymptotically large number of uses and vanishing error (previously defined in [4]). Note that these unitaries can communicate in either direction, or both simultaneously. We are primarily interested in the combined rate in both directions (as with simulation costs). Let $C^{\mathrm{ent}}_{\mathrm{cap},\epsilon}(U)$ and $C^{\mathrm{EPR}}_{\mathrm{cap},\epsilon}(U)$ denote the largest number of bits that U can transmit *in a single use* up to error ϵ, when allowed arbitrary entanglement or free EPR pairs, respectively. The corresponding asymptotic capacities are denoted $C^{\mathrm{ent}}_{\mathrm{cap}}(U)$ and $C^{\mathrm{EPR}}_{\mathrm{cap}}(U)$. (Previous works [4, 5] used the notation $C^{E}_{+}(U)$ for the latter scenario.) Likewise, let $E_{\mathrm{cap}}(U)$ denote the asymptotic entanglement capacity. Naturally, simulation costs are upper bounds to communication capacities.

We might reasonably expect that these capacities reflect the interaction strength of the unitaries, and thus if one capacity is large, the others should be as well. For example, a gate that communicates well in the forward direction ought to also do so in the backward direction, and a highly entangling gate should also disentangle or communicate a lot. This is indeed the case for some well-studied unitaries (e.g., CNOT, SWAP, and unitaries close to the identity). Additionally, it has been proven that if one of these capacities is positive, the others are as well [4], and that communication capacities are generally lower bounds of the entanglement capacity ($C^{\mathrm{ent}}_{\mathrm{cap}}(U) = C^{\mathrm{EPR}}_{\mathrm{cap}}(U) \leq E_{\mathrm{cap}}(U) + E_{\mathrm{cap}}(U^\dagger)$) [4, 6]. However, beyond the above proven bounds, little support was found for the intuition. More recently, Ref. [5] finds gates exhibiting arbitrarily large differences between entanglement and disentanglement capacities, (see also [7]), and between forward and backward communication ca-

pacities. In this paper, we demonstrate the remaining separation: an arbitrarily large difference between entanglement capacity and communication capacity. Together with the results of [5], this indicates that most unitary gate capacities of interest can vary nearly independently.

The gate U. For our gate U, A and B each have $d{+}1$ dimensions (or equivalently, $n = \log(d{+}1)$ qubits) and a basis given by $\{|0\rangle, \cdots, |d\rangle\}$. Let $|\Phi\rangle = \frac{1}{\sqrt{d}}\,(\,|11\rangle + \cdots + |dd\rangle\,)$ and $P = |00\rangle\langle 00| + |\Phi\rangle\langle\Phi|$. Define

$$U = |00\rangle\langle\Phi| + |\Phi\rangle\langle 00| + I - P.$$

In other words, U swaps $|00\rangle$ with $|\Phi\rangle$ and leaves the rest of the space (i.e. the support of $I - P$) unchanged. Note that $U = U^\dagger$.

We consider this gate U because it can certainly create or remove $\log d \approx n$ ebits but it leaves most of the space unchanged. This latter property will allow us to simulate U with little communication, implying upper bounds on its communication capacity.

The simulation protocol W. Define $|\phi_-\rangle = \frac{1}{\sqrt{2}}\,(\,|\Phi\rangle - |00\rangle\,)$. Note that U has only 1 nontrivial eigenvalue, -1, and the corresponding eigenvector is $|\phi_-\rangle$. Let $\mathcal{M}_{\mathrm{i}}$ be the ideal coherent measurement that maps $|\phi_-\rangle|0\rangle \to |\phi_-\rangle|0\rangle$ and $|\phi\rangle|0\rangle \to |\phi\rangle|1\rangle$ if $\langle\phi|\phi_-\rangle = 0$. $\mathcal{M}_{\mathrm{i}}$ is a 2-outcome measurement with POVM elements $M_0 = |\phi_-\rangle\langle\phi_-|, M_1 = I - |\phi_-\rangle\langle\phi_-|$. The protocol W simulates U by using a nonlocal state identification procedure $\mathcal{M}_{\mathrm{a}}$ (described below) that will make use of $|\phi_-\rangle^{\otimes m-1}$ to approximate $\mathcal{M}_{\mathrm{i}}$. W has 5 steps:

1. Adjoin ancillas $|\phi_-\rangle^{\otimes m-1}$.
2. Apply $\mathcal{M}_{\mathrm{a}}$. Store the outcome 0/1 in a qubit C in Bob's possession (WLOG). We will prove later that $\mathcal{M}_{\mathrm{a}}$ differs from $\mathcal{M}_{\mathrm{i}}$ in the diamond norm [8] by no more than $O(m^{-1/2})$ using the catalyst $|\phi_-\rangle^{\otimes m-1}$ and $\log(m)$ qubits of communication in each direction.
3. Apply the gate $\mathrm{Diag}(-1, 1)$ to C, so that $|0\rangle$ is mapped to $-|0\rangle$ and $|1\rangle$ mapped to $|1\rangle$.
4. Reverse $\mathcal{M}_{\mathrm{a}}$ in step 1, so as to coherently erase the outcome in C. This step also requires $\log(m)$ qubits of communication in each direction.
5. Discard the ancillas and system C.

Procedure for nonlocal state identification $\mathcal{M}_{\mathrm{a}}$. We start with an informal description of the task, ignoring locality constraints. Suppose we want to know whether or not an unknown incoming state $|\beta\rangle$ is equal to some other state $|\alpha\rangle$, and we have possession of $m{-}1$ copies of $|\alpha\rangle$. One (approximate) method is to project $|\alpha\rangle^{\otimes m-1}|\beta\rangle$ onto the symmetric subspace of $(\mathbb{C}^d)^{\otimes m}$ (defined as the span of all vectors of the form $|\psi\rangle^{\otimes m}$ for $|\psi\rangle \in \mathbb{C}^d$). This defines a two-outcome measurement with measurement operators $\Pi_{\mathrm{sym}} := \frac{1}{m!}\sum_{\pi\in\mathcal{S}_m}\pi$, and $I - \Pi_{\mathrm{sym}}$. (Here $\mathcal{S}_m$ is the group of operators that permute the m registers.) The outcome corresponding to Π_{sym} occurs with probability $\langle\alpha|^{\otimes m-1}\langle\beta|\,\frac{1}{m!}\sum_{\pi\in\mathcal{S}_m}\pi\,|\alpha\rangle^{\otimes m-1}|\beta\rangle$. A fraction $\frac{1}{m}$ of the permutations fix the m^{th} register. For each such π, $\langle\alpha|^{\otimes m-1}\langle\beta|\pi|\alpha\rangle^{\otimes m-1}|\beta\rangle = 1$. The remaining $1{-}\frac{1}{m}$ fraction of the permutations swaps the m^{th} register with one of the others. In this case $\langle\alpha|^{\otimes m-1}\langle\beta|\pi|\alpha\rangle^{\otimes m-1}|\beta\rangle = |\langle\alpha|\beta\rangle|^2$. Thus the probability of obtaining Π_{sym} is $\frac{1}{m} + (1{-}\frac{1}{m})|\langle\alpha|\beta\rangle|^2 = |\langle\alpha|\beta\rangle|^2 + \frac{1}{m}(1{-}|\langle\alpha|\beta\rangle|^2)$, and the procedure simulates the measurement with operators $\{|\alpha\rangle\langle\alpha|, I - |\alpha\rangle\langle\alpha|\}$ up to error at most $1/m$.

Observe that instead of π ranging over all $m!$ permutations, it would suffice to take only the m cyclic permutations. For the multi-partite setting, this will allow us to save dramatically on communication. We now describe the bipartite protocol and derive a careful bound on the accuracy.

Let $|s\rangle = \frac{1}{\sqrt{m}}\sum_{j=0}^{m-1}|j\rangle$ and S be a register prepared in the state $|s\rangle$. Let Y act on $S \otimes (\mathbb{C}^d)^{\otimes m}$ by mapping $|j\rangle|\psi_1\rangle|\psi_2\rangle\cdots|\psi_m\rangle$ to $|j\rangle|\psi_{1-j}\rangle|\psi_{2-j}\rangle\cdots|\psi_{m-j}\rangle$, with arithmetic done mod m. That is, S controls a cyclic permutation of the m registers, taking the first register to the j^{th} one if the state of S is $|j{-}1\rangle$.

With a slight abuse of notation, let $\mathcal{M}_{\mathrm{i}}$ and $\mathcal{M}_{\mathrm{a}}$ be the ideal and approximate coherent state identification protocols for some bipartite state $|\alpha\rangle$, with the answer residing with Bob. The state to be measured lives in systems AB. Alice and Bob already share $|\alpha\rangle^{\otimes m-1}$ in $A_2B_2 \otimes \cdots \otimes A_mB_m$. $\mathcal{M}_{\mathrm{a}}$ is given by:

1. Alice prepares a register S in the state $|s\rangle$.
2. Alice applies Y on $S \otimes A \otimes A_2 \cdots A_m$ (i.e. she applies the S-controlled cyclic permutation on her halves of the m bipartite systems).
3. Alice sends S to Bob using $\log(m)$ qubits of forward communication.
4. Bob performs Y on $S \otimes B \otimes B_2 \cdots B_m$ thereby completing the S-controlled cyclic permutation on the m bipartite systems.
5. Bob coherently measures S with POVM $\{|s\rangle\langle s|, I - |s\rangle\langle s|\}$. The final outcome is written to a register C in Bob's possession.
6. Bob performs $Y^\dagger$ on $S \otimes B \otimes B_2 \cdots B_m$.
7. Bob sends S to Alice using $\log(m)$ qubits of backward communication.
8. Alice applies $Y^\dagger$ on $S \otimes A \otimes A_2 \cdots A_m$.

We now show that $\mathcal{M}_{\mathrm{a}}$ approximates $\mathcal{M}_{\mathrm{i}}$ in the following sense. The diamond-norm of a superoperator $\mathcal{S}$ is defined as $\|\mathcal{S}\|_\diamond := \max_{\psi\geq 0, \mathrm{tr}\psi=1}\|(\mathcal{I}\otimes\mathcal{S})(\psi)\|_1$. We will show that $\|\mathcal{M}_{\mathrm{a}} - \mathcal{M}_{\mathrm{i}}\|_\diamond \leq \frac{\sqrt{2}}{\sqrt{m}}$. Consider the state $|\phi\rangle = \sqrt{p}\,|a_0\rangle_R|\alpha\rangle_{AB} + \sqrt{1{-}p}\,|a_1\rangle_R|\alpha_\perp\rangle_{AB}$, where R is

a reference system that may be entangled with the incoming systems AB, $\langle\alpha_\perp|\alpha\rangle_{AB} = 0$, and $|a_0\rangle, |a_1\rangle$ are unit vectors that are not necessarily orthogonal to one another. This is the most general initial state. Evolving $|\phi\rangle$ according to $\mathcal{M}_{\rm a}$ gives a final state $|{\rm fin}\rangle =$

$$\sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle|0\rangle + \sqrt{1-p}\,|a_1\rangle|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1}|s\rangle|1\rangle + |{\rm err}\rangle$$

where $|{\rm err}\rangle =$

$$\tfrac{\sqrt{2}}{m^{3/2}}\sum_{jj'}\sqrt{1-p}\;|a_1\rangle|\alpha\rangle^{\otimes j-j'}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-(j-j')}|j'\rangle|-\rangle$$

and $|-\rangle = \frac{1}{\sqrt{2}}(\,|0\rangle - |1\rangle\,)$. The first two terms in $|{\rm fin}\rangle$ are precisely the state $|{\rm cor}\rangle$ obtained by applying $\mathcal{M}_{\rm i}$ to $|\phi\rangle$. The last term $|{\rm err}\rangle$ represents the deviation. The derivation is routine and is deferred to the appendix. When calculating $|\langle{\rm cor}|{\rm err}\rangle|$, only terms with $j = j'$ contribute to the inner product. There are m such terms, all being the same, giving the bound $|\langle{\rm cor}|{\rm err}\rangle| \le \frac{\sqrt{1-p}}{\sqrt{m}}$ and matching precisely the probability of failure given by the informal argument. It also gives $|\langle{\rm cor}|{\rm fin}\rangle| \ge 1 - \frac{\sqrt{1-p}}{\sqrt{m}} \ge 1 - \frac{1}{\sqrt{m}}$. We are now ready to apply the well known relation

$$\tfrac{1}{2}\|\,|a\rangle\langle a| - |b\rangle\langle b|\,\|_1 = \sqrt{1-|\langle a|b\rangle|^2} \le \sqrt{2\,(1-|\langle a|b\rangle|)}$$

to bound $\|\mathcal{M}_{\rm a} - \mathcal{M}_{\rm i}\|_\diamond$ which is equal to

$$\begin{aligned} &= \sup_{|\phi\rangle}\|(\mathcal{I}\otimes\mathcal{M}_{\rm a})(|\phi\rangle\langle\phi|) - (\mathcal{I}\otimes\mathcal{M}_{\rm i})(|\phi\rangle\langle\phi|)\|_1 \\ &= \sup_{|\phi\rangle}\|\,|{\rm cor}\rangle\langle{\rm cor}| - |{\rm fin}\rangle\langle{\rm fin}|\,\|_1 \le \tfrac{\sqrt{2}}{\sqrt{m}}\,. \end{aligned}$$

Returning to the protocol W that simulates U, if we replace the two uses of $\mathcal{M}_{\rm a}$ by $\mathcal{M}_{\rm i}$, we obtain an exact implementaion of U. By the triangle inequality, $\|U-W\|_\diamond \le 2\|\mathcal{M}_{\rm a} - \mathcal{M}_{\rm i}\|_\diamond \le \frac{2\sqrt{2}}{\sqrt{m}}$. For W to simulate U with accuracy ϵ, it suffices to take $m = \frac{8}{\epsilon^2}$. The simulation consumes $2\log m$ qubits of communication in each direction. Thus we have the following.

Theorem 1 $C^{\rm ent}_{{\rm sim},\epsilon}(U) \le 24 + 16\log\frac{1}{\epsilon}$.

Here U is implicitly parameterized by the system size n, yet the simulation cost is independent of it. Note as well that the nonlocal state identification protocol generalizes straightforwardly to more than two remote parties (say, k). One way to do this is for one party to create $|s\rangle$ which is then circulated among all parties. Another way is to have the k parties sharing $|s\rangle = \frac{1}{\sqrt{m}}\sum_{j=0}^{m-1}|j\rangle^{\otimes k}$, each sends his share to the party designated to have the answer, and has the share returned to complete the protocol. Next we prove two results based on the simulation protocols and Theorem 1.

Consequence 1: EPR pairs are suboptimal for simulation cost.

Let $\delta := \sqrt[8]{4\epsilon}$. We claim that

$$C^{\rm EPR}_{{\rm sim},\epsilon}(U) \ge \Delta_\epsilon := 2\log(d) - 1 + \log((1-2\delta)(1-\delta)^2)\,. \quad (1)$$

Proof. Let $|\varphi\rangle = \frac{1}{\sqrt{2}}(|\Phi\rangle_{AB}\otimes|00\rangle_{A'B'} + |00\rangle_{AB}\otimes|\Phi\rangle_{A'B'})$. We consider the state-change from $|\varphi\rangle$ to $U_{AB}\otimes I_{A'B'}|\varphi\rangle$. Recall that $|\Phi\rangle = \frac{1}{\sqrt{d}}(|11\rangle + \cdots + |dd\rangle)$, so $|\varphi\rangle$ is just a maximally entangled state of Schmidt rank $2d$. By Corollary 10 of Ref [9], preparing $U_{AB}\otimes I_{A'B'}|\varphi\rangle$ up to fidelity $1-\epsilon$ from any maximally entangled state requires an amount of communication at least as large as Δ_ϵ. Therefore, simulating U up to error ϵ given unlimited EPR pairs requires Δ_ϵ bits of communication, contrary to the $O(\log\frac{1}{\epsilon})$ bits of communication in Theorem 1 when $O(\frac{1}{\epsilon^2})$ copies of $|\phi_-\rangle$ are available. □

Note that *any* $n\times n$-qubit unitary can be trivially simulated with EPR pairs and $4n$ bits of communication by teleporting Alice's input to Bob, having him apply the unitary and then teleporting her system back. Thus, Eq. (1) implies that even given unlimited EPR pairs and allowing a small error, simulating U is at least half as costly as simulating a completely general unitary on $n\times n$ qubits.

Consequence 2: Some gates can entangle exponentially more than they can communicate.

Since $U|00\rangle = |\Phi\rangle$, we can bound $E_{\rm cap}(U) \ge \log(2^n - 1) \approx n$. On the other hand, we have:

Theorem 2 *For any $c > 2$ and for all n sufficiently large, $C^{\rm ent}_{\rm cap}(U) \le 8c\log n$.*

When communicating using a gate in both directions simultaneously, there is generally a tradeoff between the forward and bacward communication rates. The one-way capacity in each direction is an extreme point of that tradeoff. We denote these capacities by $C^{\rm ent}_{{\rm cap},\rightarrow}(U)$ and $C^{\rm ent}_{{\rm cap},\leftarrow}(U)$. Theorem 2 can be proved by showing $C^{\rm ent}_{{\rm cap},\rightarrow}(U) \le 4c\log n$, since the symmetry of U means that the same bound applies to $C^{\rm ent}_{{\rm cap},\leftarrow}(U)$, and finally we can bound $C^{\rm ent}_{\rm cap}(U) \le C^{\rm ent}_{{\rm cap},\rightarrow}(U) + C^{\rm ent}_{{\rm cap},\leftarrow}(U) \le 8c\log n$.

Proof of $C^{\rm ent}_{{\rm cap},\rightarrow}(U) \le 4c\log n$.

The nonlocal state identification protocol $\mathcal{M}_{\rm a}$ uses shared entangled states between Alice and Bob and $\log m$ qubits of communication in each direction, and the protocol W that simulates U uses $\mathcal{M}_{\rm a}$ twice, W uses $2\log m$ qubits of forward communication. But back communication and shared entanglement cannot increase the classical capacity of a noiseless forward quantum channel beyond the superdense coding bound [10], thus

$$C^{\rm ent}_{{\rm cap},\rightarrow}(W) \le 4\log m\,. \quad (2)$$

It remains to show that $C^{\rm ent}_{{\rm cap},\rightarrow}(W) \approx C^{\rm ent}_{{\rm cap},\rightarrow}(U)$ if $\|W - U\|_\diamond$ is small. To make this quantitative, we prove the following *continuity bound* in the appendix.

Lemma 3 *If $\mathcal{N}_1, \mathcal{N}_2$ are bidirectional channels with outputs in $\mathbb{C}^{d+1}\otimes\mathbb{C}^{d+1}$ such that $\|\mathcal{N}_1 - \mathcal{N}_2\|_\diamond \le \epsilon$, then*

$$|C^{\rm ent}_{{\rm cap},\rightarrow}(\mathcal{N}_1) - C^{\rm ent}_{{\rm cap},\rightarrow}(\mathcal{N}_2)| \le 8\epsilon\log(d+1) + 4H_2(\epsilon)$$

where H_2 is the binary entropy function.

Our continuity bound means that the more accurate $\mathcal{M}_{\rm a}$ is, the closer the capacities of U and W are. On the other hand, making $\mathcal{M}_{\rm a}$ more accurate requires more communication. Thus we face a trade-off between keeping the capacity of W small and keeping the capacities of U and W close to each other. Optimizing will give us a bound of $O(\log n)$ bits on the capacity of U.

Completing the proof of $C^{\rm ent}_{{\rm cap},\rightarrow}(U) \leq 4c \log n$.

Recall that the accuracy of the approximate nonlocal state identification in terms of the communication cost is $\eta = \frac{\sqrt{2}}{\sqrt{m}}$, and that $\|U-W\|_\diamond \leq 2\eta = \epsilon$. According to Lemma 3, since $\log(d+1) = n$, the difference in the capacities of U and W is suppressed if $m = n^c$ for $c > 2$. More precisely,

$$\begin{aligned} C^{\rm ent}_{{\rm cap},\rightarrow}(U) &\leq C^{\rm ent}_{{\rm cap},\rightarrow}(W) + 16\eta \log(d+1) + 4H_2(2\eta) \\ &\leq 4\log m + 16\eta n + 4\sqrt{2\eta} \\ &\leq 4c \log n + 16\sqrt{2} n^{1-c/2} + 8 \cdot 2^{0.75} n^{-c/4} \end{aligned}$$

where each term is bounded by the corresponding term in the subsequent line (and $H_2(x) \leq 2\sqrt{x}$). □

Extensions. Our simulation procedure allows us to simulate any bipartite gate with r non-trivial eigenvalues using $O(r \log(r/\epsilon))$ qubits of communication. This is accomplished by testing the state held by Alice and Bob sequentially against each of the r corresponding eigenvectors. Each individual test needs to have error ϵ/r so that the total error can be bounded by ϵ. This simulation method is useful for $r \ll \log(d)$ (since a gate can be trivially simulated using $\log d$ qubits of communication in each direction).

Regarding unitary gate capacities, we have shown that $C^{\rm ent}_{\rm cap}(U)$ can scale like the logarithm of $E_{\rm cap}(U)$. However, it is unknown how much further this result could be improved. For our example, it is possible that $C^{\rm ent}_{\rm cap}(U)$ can be upper-bounded by a constant even as $n \rightarrow \infty$. Moreover, it is possible that even stronger separations are possible. Bound 1 of [4] implies that $C^{\rm ent}_{\rm cap}(U) > 0$ whenever $E_{\rm cap}(U) > 0$, but even for fixed dimension no nonzero lower bound on $C^{\rm ent}_{\rm cap}(U)$ is known. The difficulty is that the proof in [4] relates $C^{\rm ent}_{\rm cap}(U)$ to the amount of entanglement which one use of U can create from unentangled inputs. This quantity can be arbitrarily smaller than $E_{\rm cap}(U)$ even for fixed dimensions.

Acknowledgements. We are grateful to Charles Bennett, whose hope for a simple theory concerning interconversions between nonlocal resources has inspired many of our investigations in this subject.

AWH was funded by the U.S. ARO under grant W9111NF-05-1-0294, the EC under Marie Curie grants ASTQIT (FP6-022194) and QAP (IST-2005-15848), and the U.K. EPSRC through "QIP IRC." DWL was funded by the CRC, ORF, NSERC, CIFAR, MITACS, and QuantumWorks.

[1] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels", Phys. Rev. Lett., **70** (1993) p. 1895-98.

[2] C. H. Bennett and S. J. Wiesner, "Communication via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States", Phys. Rev. Lett, **69** (1992) p. 2881-2884.

[3] C.H. Bennett, H.J. Bernstein, S. Popescu and B. Schumacher. "Concentrating entanglement by local operations." Phys. Rev. A **53** 2046–2052 (1996), arXiv:quant-ph/9511030.

[4] C. H. Bennett, A. W. Harrow, D. W. Leung, and J. A. Smolin, "On the capacities of bipartite Hamiltonians and unitary gates," IEEE Trans. Inf. Theory **49**, 1895 (2003), arXiv:quant-ph/0205057.

[5] A. W. Harrow and P. W. Shor, "Time reversal and exchange symmetries of unitary gate capacities," arXiv:quant-ph/0511219v1.

[6] D. W. Berry and B. C. Sanders, "Relation between classical communication capacity and entanglement capability for two-qubit unitary operations," Phys. Rev. A **68**, 032312 (2003), arXiv:quant-ph/0207065.

[7] N. Linden, J. A. Smolin, and A. Winter, "The entangling and disentangling power of unitary transformations are unequal," Phys. Rev. Lett. **103** 030501 (2009). arXiv:quant-ph/0511217.

[8] A. Kitaev, A. Shen, and M. Vyalyi, "Classical and Quantum Computation," American Mathematics Society Press (2000).

[9] P. Hayden and A. J. Winter, "On the communication cost of entanglement transformations," Phys. Rev. A, **67**, 012306 (2003), arXiv:quant-ph/0204092.

[10] R. Cleve, W. van Dam, M. A. Nielsen, and A. Tapp, "Quantum entanglement and the communication complexity of the inner product function," *Proceedings of the 1st NASA International Conference on Quantum Computing and Quantum Communications, Lecture Notes in Computer Science*. 1998, vol. 1509, pp. 61–74, Springer-Verlag, arXiv:quant-ph/9708019.

[11] A. M. Childs, D. W, Leung, and H.-K. Lo, "Two-way quantum communication channels," Int. J. Quant. Inf., **4** 63-83 (2006), arXiv:quant-ph/0506039.

[12] A. S. Holevo, "Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel," Problems of Information Transmission, **9** 177-183 (1973).

[13] M. Fannes, "A continuity property of the entropy density for spin lattice systems," Comm. Math. Phys. **31** 291 (1973).

[14] R. Alicki and M. Fannes, "Continuity of quantum mutual information," J. Phys. A, **37** L55 (2004), arXiv:quant-ph/0312081.

PROOFS

Deriving the state evolved by $\mathcal{M}_{\rm a}$

We use all the notations defined in the main text. In the proof of $\|\mathcal{M}_{\rm a}-\mathcal{M}_{\rm i}\|_\diamond \le \frac{\sqrt{2}}{\sqrt{m}}$, we claim that the output state of applying $\mathcal{M}_{\rm a}$ to the most general initial state $|\phi\rangle = \sqrt{p}\,|a_0\rangle_R|\alpha\rangle_{AB} + \sqrt{1-p}\,|a_1\rangle_R|\alpha_\perp\rangle_{AB}$ is of a certain form. Here is a justification of this fact. The state after attaching the ancillas is:

$$\sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle + \sqrt{1-p}\,|a_1\rangle|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1}|s\rangle\,.$$

After Alice applies Y, communicates S to Bob, and Bob applies Y:

$$\sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle + \sqrt{1-p}\,|a_1\rangle\tfrac{1}{\sqrt{m}}\sum_j |\alpha\rangle^{\otimes j}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-j}|j\rangle\,.$$

Now Bob attaches $|0\rangle_C$ and makes the coherent measurement on S, taking $|s\rangle|0\rangle_C \to |s\rangle|0\rangle_C$ and $|s_\perp\rangle|0\rangle_C \to |s_\perp\rangle|1\rangle_C$ for all $\langle s_\perp|s\rangle = 0$. To write down the resulting state, we should rewrite each $|j\rangle$ in the Fourier basis which includes $|s\rangle$. But to obtain just a bound, we can simply express $|j\rangle = \frac{1}{\sqrt{m}}|s\rangle + \frac{\sqrt{m-1}}{\sqrt{m}}|s_j\rangle$ where $\langle s_j|s\rangle = 0$. The measurement on S thus results in the state

$$\sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle|0\rangle + \sqrt{1-p}\,|a_1\rangle\tfrac{1}{\sqrt{m}}\sum_j |\alpha\rangle^{\otimes j}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-j}\left(\,\tfrac{1}{\sqrt{m}}|s\rangle|0\rangle + \tfrac{\sqrt{m-1}}{\sqrt{m}}|s_j\rangle|1\rangle\,\right).$$

Here, the second occurrence of the $|s\rangle|0\rangle$ term (the one in the parenthesis) represents an erroneous measurement outcome. We add and subtract $\frac{1}{\sqrt{m}}|s\rangle|1\rangle$ in the parenthesis:

$$\sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle|0\rangle + \sqrt{1-p}\,|a_1\rangle\tfrac{1}{\sqrt{m}}\sum_j |\alpha\rangle^{\otimes j}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-j}\left(\,\tfrac{1}{\sqrt{m}}|s\rangle(|0\rangle-|1\rangle) + |j\rangle|1\rangle\,\right).$$

Rearranging, we get:

$$\sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle|0\rangle + \sqrt{1-p}\,|a_1\rangle\tfrac{1}{\sqrt{m}}\sum_j |\alpha\rangle^{\otimes j}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-j}|j\rangle|1\rangle$$

$$+\sqrt{1-p}\,|a_1\rangle\tfrac{1}{\sqrt{m}}\sum_j |\alpha\rangle^{\otimes j}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-j}\tfrac{\sqrt{2}}{\sqrt{m}}|s\rangle|-\rangle$$

where the first line is what an ideal measurement will produce (with unit norm), and the second line represents an error term (and it is *not* orthogonal to the ideal portion, since the sum is also normalized). Now, Bob applies $Y^\dagger$ and sends S back to Alice, who then applies $Y^\dagger$, resulting in the final state $|\text{fin}\rangle = |\text{cor}\rangle + |\text{err}\rangle$ where

$$|\text{cor}\rangle = \sqrt{p}\,|a_0\rangle|\alpha\rangle^{\otimes m}|s\rangle|0\rangle_C + \sqrt{1-p}\,|a_1\rangle|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1}|s\rangle|1\rangle_C$$

$$|\text{err}\rangle = \tfrac{\sqrt{2}}{m^{3/2}}\sum_{jj'}\sqrt{1-p}\,|a_1\rangle|\alpha\rangle^{\otimes j-j'}|\alpha_\perp\rangle|\alpha\rangle^{\otimes m-1-(j-j')}|j'\rangle|-\rangle_C$$

as claimed.

We can bound $\||\text{err}\rangle\|_2$ by inspecting the expression right after the rearrangement, which gives $\||\text{err}\rangle\|_2 \le \frac{\sqrt{2(1-p)}}{\sqrt{m}}$. This implies $|\langle\text{cor}|\text{fin}\rangle| \ge 1 - |\langle\text{cor}|\text{err}\rangle| \ge 1 - \frac{\sqrt{2(1-p)}}{\sqrt{m}}$. Alternatively, we can explicitly calculate $|\langle\text{cor}|\text{err}\rangle|$ using their expressions given above. Only the $j = j'$ terms contribute to the inner product. But there are m such terms, all being the same, giving the slightly better bound $|\langle\text{cor}|\text{err}\rangle| \le \frac{\sqrt{1-p}}{\sqrt{m}}$ and matching the probability of failure given by the informal argument.

Proof of Lemma 3: Our proof will closely parallel that of Lemma 1 of [5], which is similar to the above but holds for the case when $\mathcal{N}_1$ and $\mathcal{N}_2$ are isometries. The main ingredient in both proofs is a single-shot capacity formula for bidirectional channels, first established for isometries in [4], but then extended to arbitary bidirectional channels in [11]:

$$C^{\text{ent}}_{\text{cap},\rightarrow}(W) = \sup_{\rho^{XAA'BB'}} I(X;BB')_{W(\rho)} - I(X;BB')_{\rho}. \tag{3}$$

Here A, B are the registers acted on by W, A', B' are ancillas of arbitrary dimension, X is a classical register, $I(X;Y) = H(X) + H(Y) - H(XY)$ is the quantum mutual information of the state given by the subscript. $H(R) = H(\sigma) = -\text{tr}\sigma\log\sigma$ is the von Neumann entropy for the reduced density matrix σ on the system R. When one of the registers X is classical, the state on XY represents an ensemble of quantum states on Y labeled by basis states of X, and the quantum mutual information is the Holevo information [12]. Eq. (3) can be interpreted to mean that $C^{\text{ent}}_{\text{cap},\rightarrow}(W)$ equals the largest single-shot increase in mutual information possible when applying W to any ensemble of bipartite states. Due to Eq. (3),

$$C^{\text{ent}}_{\text{cap},\rightarrow}(U) - C^{\text{ent}}_{\text{cap},\rightarrow}(W) \leq I(X;BB')_{U(\rho)} - I(X;BB')_{W(\rho)} \tag{4}$$

where ρ attains the supremum in the expression for $C^{\text{ent}}_{\text{cap},\rightarrow}(U)$ to some arbitrary precision. (This precision parameter is independent from all other parameters considered, and thus will be omitted for simplicity.)

Thus the desired continuity bound is essentially a continuity result for quantum mutual information. The crucial challenge is the lack of dimensional bounds on the systems X and B', so that Fannes inequality [13] does not provide the needed continuity result. Instead, we use a generalization due to Fannes and Alicki [14] that applies to conditional entropy:

$$|H(Y|Z)_{\sigma} - H(Y|Z)_{\sigma'}| \leq 4\epsilon\log d + 2H_2(\epsilon)\,,$$

where $\epsilon = \|\sigma - \sigma'\|_1$ and $d = \dim Y$. Remarkably, this Fannes-Alicki inequality provides an upper bound that is independent of the size of the conditioned system Z.

Returning to Eq. (4), first note that if $\|W - U\|_{\diamond} \leq \epsilon$, then $\|W(\rho) - U(\rho)\|_1 \leq \epsilon$. Next, we can expand $I(X;BB')$ as

$$I(X;BB') = H(B') + H(B|B') - H(B|B'X) - H(B'|X)\,.$$

We now bound the difference of each of the above terms when evaluated on $W(\rho)$ and $U(\rho)$. The $H(B')$ and $H(B'|X)$ terms are the same for both states since W and U act only on A, B. Applying the Fannes-Alicki inequality to the remaining two terms and using $\dim B = d + 1$ establishes the Lemma.